DOCUMENT RESUME

ED 393 889 TM 024 810

AUTHOR Keyes, Marie; And Others
TITLE Performance Assessment: Mississippi at the Cusp.
PUB DATE 8 Nov 95
NOTE 26p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Biloxi, MS, November 8-10, 1995).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Administrator Attitudes; Administrators; Difficulty Level; *Educational Assessment; Elementary School Students; *Elementary School Teachers; Elementary Secondary Education; Graduate Study; Inservice Teacher Education; *Performance Based Assessment; Professional Development; Secondary School Students; Secondary School Teachers; State Programs; *Student Attitudes; Surveys; *Teacher Attitudes; Testing Programs; *Test Use
IDENTIFIERS Large Scale Assessment; *Mississippi

ABSTRACT

This paper documents student, teacher, and administrator attitudes toward the initiation of performance-based tests on a statewide level in Mississippi. The initial study surveyed approximately 620 test participants in grades 1 through 10 during the period between the fall test administration and the receipt of the test results in January. The teacher survey (n=58) indicated that teachers saw the tests as more difficult than did students, but their overall perceptions were more positive. Administrators from 11 of 14 schools returned their surveys, and their responses were generally positive. The ongoing study, conducted by educators, examines the direction Mississippi appears to be taking. Vital to the preparedness of the state's teachers for performance assessment is a return to the classroom as students. Increased access to graduate educational literature, keeping up to date with current learning theories, and interactive work with peers will assure that performance-based assessment will have a bright future in Mississippi. An attachment describes the participant study in detail. (Contains 5 references in the overview, 15 in the attached report, and 2 figures.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERFORMANCE ASSESSMENT:

MISSISSIPPI AT THE CUSP

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
FELICIA ROBINSON
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ED 393 889

Presentation November 8, 1995 in Biloxi, MS at the

Mid South Educational Research Association Annual Meeting

Presenters:

Marie Keyes
Powers Elementary School, Jones County Public Schools,

Felicia Robinson
Mendenhall High School, Simpson County School District

Linda Walters
Hattiesburg High School, Hattiesburg Public Schools

Written in conjunction with
Read M. Diket, Ph.D.
William Carey College

TM024810

BEST COPY AVAILABLE

Performance Assessment: Mississippi at the Cusp

Marie Keyes, Felicia Robinson, and Linda Walters

with Read M. Diket

Presentation November 8, 1995, in Biloxi, Mississippi

Mid-South Educational Research Association Annual Meeting

Abstract: This paper documents student, teacher, and administrator attitudes toward the initiation of performance based tests on a statewide level in Mississippi. The initial study surveyed approximately 620 test participants during the period between the fall test administration and receipt of the test results in January 1995. The ongoing study, conducted by educators, examines the direction Mississippi appears to be taking.

In the fall of 1994, Mississippi initiated a combination performance based assessment (PBA) and norm referenced test (NRT). Performance assessment gives educators a clearer indication of the student's ability; indeed, a performance assessment combined with norm-referenced testing (as in Mississippi) provides for a more comprehensive evaluation than either test provides singularly. Performance based testing moves beyond rote memorization and mechanical computations. Open-ended questions require the student to demonstrate problem solving, analytical, and written communication skills.

Hymes (1990) proposes that performance assessment goes beyond rote memorization toward demonstrated understanding of concepts and issues. Public schools in several states have instituted performance based assessment and the majority report a better "read" on student performance than with norm-referenced testing. State use of performance results is relatively new, thus most states have published only preliminary results.

The majority of these states report a positive attitude towards PBA. Mississippi survey responses appear to be in line with other states. Most respondents indicate a positive attitude in regard to new testing procedures and display little negativity concerning testing conditions.

Student surveys, consisting of ten questions, measured student response on a three point Likert scale. Five student questions paralleled the teacher surveys. The remaining questions on the teacher survey questioned educators about diagnostic uses of test results and expectations for instructional change. Administrator surveys asked about the preparation of personnel who administered the test and sought expectations about test results relative to the goals and objectives of schools.

As a result of the data, the researchers devised the following action plan:

+ Discussion with students will change from meeting grade assignment to meeting and exceeding criteria.
+ Teachers will move toward embedding skills in context, crossing discipline areas.
+ Teachers will recognize the many genres in which students can perform (i.e. writing essays or making graphs).
+ Teachers will expect students to defend their answers.

Test results were received in Mississippi schools in late January and early February 1995. After further study, researchers found that many educators did not wait for the test results to implement what they felt were the needed steps to improvement. Many of the educators interviewed indicated the urgent need for fundamental changes in classroom instruction in order to improve students' performance and preparedness on the PBA.

Performance Based Assessment: An Elementary School's Response

Expectations for change in educational practices were high with the integration of performance assessment in the fall of 1994 for the state of Mississippi. The examination of one participant elementary school since the publication of "Fall 1994 Performance Based Assessment Study of Participants' Attitudes" (Diket, Robinson, et al, 1995) reflects administrator, teacher and superintendent responses to a change in evaluation methods. Results of the PBA innovation provided educators with strengths and weaknesses of school related content areas as well as outcomes of present instructional methods. Test results received in January provided necessary information for planning reform of the educational system.

The participant elementary school from the fall survey has yet to initiate change in methods and procedures in response to 1994 test outcomes. Surprisingly, teachers continue to teach to the Stanford Achievement Test format by allowing limited question response on evaluation criteria. Students are not expected to justify answers; rather they select from multiple choices. Curricula changes are limited to new mathematics programs in most, but not all, grade levels. The incorporation of creative writing coursework in grades four through seven provides students with instruction on critical thinking disciplines and self expression. Cooperative learning instruction is limited, with only one course in grades six and seventh actively utiizing the technique. Lower grades, one through four, show little or no change in teaching methods as a result of the performance based assessment.

The lack of change in this participant elementary school suggests the need to emphasize the importance of results and consequences of not responding to critical information. Teachers need extensive training on new teaching strategies. A majority of the teachers at the participating school have not returned to college coursework since their graduation. In-service programs so far fail to emphasize the importance of test outcomes and fail to identify teacher modifications that need implementation. Communication between district officials and the local school are minimal as to testing outcomes.

Change in educational systems, historically, has been a slow process. The classroom is the starting point for all educational reform. Reaching the classroom appears to be a critical problem for this participant elementary school. It will take several years to implement changes in methods and curricula. The performance based assessment should alert students to a higher level of processing and applying knowledge. Teachers must lead students to this higher level by changing and updating instruction. Portfolio examination needs to be added to classroom practices.

Conversation with a District

"Yes ma'm, I liked this test. None of those other tests ever let me explain about my answers." When the fifth grade teacher in an area school heard this, she knew that this kind of testing would give the child the power to prove that he was learning and that he understood what

he was learning. While performance testing is new to the state, in many districts the seeds for scucess had already been planted.

Some school districts, such as the Hattiesburg Public Schools, began shifting away from the facts-only presentation in favor of instructional strategies emphasizing writing throughout the curricula, critical thinking, whole language and thematic approaches, portfolios, and extensive use of manipulatives for all children. As a result of these efforts emphasizing the higher order thinking skills, the students appear to be comfortable with the performance testing.

Hattiesburg serves a population which approaches seventy-five percent economically disadvantaged, and the usual standardized test results show scores hovering in fortieth percentile. These scores would usually correlate with low performance assesment results but this is not the case in the Hattiesburg Public Schools. Performance assessment scores range from low average to high average in all categories with the scores rising consistently in the upper grades. This anomaly shows that the emphasis on teaching with performance assessment in mind has made a difference in test scores as well as attitudes.

Before the performance assessment testing in Mississippi, school districts throughout the state anticipated norm referenced testing with a combination of dread and resignation because so often the results of the standardized tests depended on the memorization skills of the students and external factors which seemed to be out of control of teachers and administrators. While the uncontrollable external factors are still present, educational professionals at all levels are finding that the performance assessments allow students to work through the reasoning and thinking process in order to come to conclusions that reveal their knowledge.

In Search of a Theory

Performance based assessment is associated with theoretical discussions of competence and performance (Shohamy, 1995). PBA evaluates educational progress which is criterion based or referenced to standards. In Mississippi's version of PBA students demonstrate in written formats their knowledge and skills in responses to test tasks. The Riverside Publishing Company, designers of the Mississippi PBA, include "tasks using open-ended format that require students to

construct a response, create a product, or perform a demonstration; focus primarily on the process, but product can also be important" (handout from Riverside, Glossary of Terms, p. 2). The Riverside PBA tasks require complex cognitive activity which shows up as multiple level responses. Scorers use a rubric (scaled lists of characteristics which describe criteria necessary to score points on each outcome). The rubric includes sample responses, "anchors" and the top "exemplar", for comparison purposes. Validity problems associated with PBA include low generalizability and sampling variability associated with tasks. Consideration of validity issues must be tied to uses and interpretations of assessment results rather than to evaluations of test instruments (Linn, 1994). Linn further maintains that "multiple types of evidence are needed in arriving at an integrated judgment regarding the validity of a particular use or interpretation" (p. 6). We infer from Linn's argument that if students' scores are to be used diagnostically (for example, to design individualized learning environments which will lead to enhanced task performance) then teachers have the best opportunity to compare personal observations and portfolios of student work to test results. With use and interpretation focus at the local level, validity problems become negligible. The value assigned to standards and expectations for "positive" consequences drive the current trend towards performance-based assessment; standards seem to have greater meaning at a state level while "positive" consequences seem more likely at the local rather than state level.

Resnick (1994) notes that performance assessments can both define standards and encourage efforts to meet those standards. However, the theoretical aspects of performance assessment need to be treated more explicitly. If performance assessment is designed with different social functions in mind as compared with traditional American testing (norm based), then appropriate epistemological assumptions need articulation. Epistemology refers to the study of the nature of knowledge and justification, more specifically to its defining features, substantiative conditions, and limits (Audi, 1995). Old, traditional arguments about knowledge and justification continue relative to analysis, sources, and validity in testing. Some models of performance assessment appear to decontextualize observed learning within a modified behaviorist theory, a view that assumes a functional understanding of mental phenomena (i.e. Riverside's test for Mississippi),

while other performance assessments contextualize learning in a way that allows situated justification against backgrounds of beliefs, varying from one context of inquiry to another (i.e. Maryland). Can we conceive of situations where both arguments have some relative merit? State-by-state comparisons subsume functional understandings; situated justifications might be best interpreted within locally held beliefs and teaching practices. Both draw from aspects of cognitive theory, pertaining to the nature of human cognition, its development and justification through processing activity. Soviet activity theory (such as is associated with Vygotsky) advances the view that cognition is both dynamic and situated in tasks which should be studied in an ecologically valid manner.

Some authors recognize that teachers are a key to a reform movement associated with performance assessment (i.e. Higuchi, 1993; Baron, 1991), most talk empowerment rather than seek the professional use of test results by teachers. Perhaps they are right-minded and wrong-headed at the same time. That is, they rightly conceptionalize teachers as a key, but think wrongly that someone external to the teachers must turn the key to unlock educational reform. In Mississippi, teachers participated in preparations for assessment reforms; their work sessions were initiated in the spirit of informed participation. In-service training prior to the October 1994 testing in their schools alerted some forward looking teachers to obvious instructional changes which would be congruent with performance assessment. Some teachers went even further; they determined ways to incorporate reform in their teaching.

Our team of teacher/researchers found some peers were already doing things "differently" in their classrooms, obviously even before assessment mandated change. These differences (observed as mid-mark norm-based test scores correlated with higher than average performance-based scores) may mark schools where constructivist (a type of reform learning theory) practices involve students as active participants in their own learning. Glaser (1992) identifies cognitive theory as the basis for innovative assessment design including the following dimensions of performance (for science): knowledge that is structured and integrated, ability to represent a problem, knowledge of procedure, automaticity (like experts), and self-regulation.

Implications

Paramount to the preparedness of Mississippi's educators is a return to the classroom as students. As graduate students, teachers learn a great deal from each other-- ideas that work, ideas that don't work, and new methods of classroom instruction. The Mississippi Institute of Higher Learning provides many of the state's public teachers with an opportunity to return to college as graduate students. In a very successful program, any Mississippi educator who wishes to achieve a Master Degree, may do so at nominal cost to that teacher. For a first Master Degree, the teacher is paid $125 per semester hour for summer course work. The requirements for acceptance are Mississippi residence and current employment with a Mississippi public elementary, middle, or secondary school. The educator may attend any Mississippi university, public or private, and these grants are available for up to five years. If the educator teaches at a Mississippi public school the academic year immediately following the summer the funds are received, then the educator does not repay the tuition. Universities in Mississippi have experienced a tremendous increase in first-time graduate school enrollment by educators. We believe this will only enhance Mississippi's assessment program.

Performance assessment, in general, relates well to standards such as modeled by GOALS 2000. Performance assessment is by nature more qualitative and situated than norm-based tests; task selection is at best problematic. Interpretations that have punitive, or negative, implications for some districts amplify the conundrum which naturally occurs with the use of performance based assessment. Teachers can help make sense of the data for their students as to test fairness, cognitive complexity, content coverage, and meaningfulness. Teacher and student responses before test results were returned indicate that they were already assuming a positively oriented processing role relative to the new test form. Mississippi provides summer funding for master level courses, thus assisting its teachers in their expanded educational role. With increased access to graduate educational literature, current learning theories, and interactive work with peers, Mississippi's teachers and their students can work towards educational reform in tandem with administrators and state officials.

References

Audi, R. (1995). *The Cambridge Dictionary of Philosophy*. Cambridge, MA: Cambridge University Press.

Diket, R., and F. Robinson, with S. Batson, D. Boyd, A. Butler, L. Callahan, G. Field, C. Garner, M. Jackson, M. Keyes, C. McLemore, G. Nickson, C. Norris, C. Parish, G. Parish, G. Thornton, and L. Walters. *Fall 1994 Performance Based Assessment: A Study of Participants' Attitudes*. Mississippi Educational Leadership, 2 (1995): 7-12.

Higuchi, C. *Performance-Based Assessments and What Teachers Need*. ERIC, 1993. TM 022 591.

Resnick, L. *Performance Puzzles: Issues in Measuring Capabilities and Certifying Accomplishments. Project 2.3, Complex Performance Assessments*. ERIC, 1994. ED 379 320.

Shohamy, W. *Performance Assessment in Language Testing*. ERIC, 1995. EJ 501 484.

Fall 1994 Performance Based Assessment:

A Study of Participants' Attitudes

Read M. Diket and Felicia Robinson coordinated this study which was undertaken by 15 teachers as a research project at William Carey College. The names of the project organizers are listed at the end of the article.

Introduction

Performance based assessments emerged in the 1990's as a viable alternative evaluation tool. Performance based assessments (PBA) require students to think critically and provide a record of processes. Admittedly, scoring is more time-consuming and expensive than with traditional, scantron-scored testing; however, resulting hard documentation of student effort provides a powerful incentive for using PBA. PBA extends educator's understandings of how their students learn and perform in school related areas.

With broad-based norm referenced testing (NRT) and PBA combinations, such as that used by Mississippi in Fall 1994, evaluations are more comprehensive. Comprehensive testing includes and moves beyond mechanical computations and fact-based, closed-ended answers. The learning and problem-solving processes document the product.

Performance assessment presents open-ended questions either orally or in writing. Students use problem solving skills to explore or analyze topics. In Hymes' (1990) estimation, performance assessment goes beyond rote memorization toward demonstrated understanding of concepts and issues. Latting (1992) also posits in his article that performance assessment represents a possible solution to problems associated with traditional testing because performance assesses direct behaviors instead of rote memorization.

Public schools in a number of states utilize performance based assessments. The majority report that performance based tests evaluate the skills of the students better than the closed-answered multiple-choice achievement test. School personnel using performance based testing acknowledge that test results will be used to improve their instructional delivery in order to prepare students for the next testing cycle.

Several states (Maryland, Vermont, and California) use performance based assessment. The tests

are relatively new, thus, most states have published only preliminary summaries. Rafferty (1993) surveyed urban teachers in Baltimore, asking them to rate Maryland's performance assessment in the second year of mandated testing. Respondents representing 66% of Baltimore's schools display a slightly positive attitude toward the tests, which varies by grade. The procedural issues Rafferty discusses are similar to those considered in this Mississippi study. Moreover, she contends that size of test groups is an important background variable as raters perceive that small groups function more smoothly on procedural tasks than larger groups.

According to Koretz (1992) the experiences to date of the Vermont portfolio program designers suggest the need for patience, moderate expectations, and ongoing formative evaluation. Neither of the basic goals of the portfolio program--to use complex performance as a measure of student performance and to use a performance-assessment program to spur instructional improvement--were obtained quickly nor easily. The Vermont experience reveals some resistance to evaluating goal outcomes through performance based assessment of portfolios.

Resistance is not the only problem associated with performance based assessments. Questions arise concerning test validity, reliability, cost, and other associated background variables. Taylor (1993) concludes that the initiative for performance based assessment should be rethought. Suggested alternatives are understanding the rationale for abandoning multiple-choice testing as mistaken, continuing the use of reformed multiple-choice testing, adopting a multiple-measures approach to assessment, and recognizing the limits of testing in accountability and educational reform. For Mississippi, multiple measures (norm referenced and performance based) may represent a justifiable cost; dual measures provide substantive information about school populations and individual students to educators who guide and evaluate educational change.

Robert Linn (1994) reports on policy promises

and technical measurement standards for performance assessment. Performance based tests and standards are being evaluated for educational worth, as are traditional standardized assessments. In the current era of change, testing and assessment issues continue to include validity, traditionally a primary concern in test evaluation. The form of assessment should not change the validity of the assessment. However, reliability appears problematic with assessments using extended, constructed responses.

The presence and the role of the federal government are much greater than seen in previous test-based accountability and reform efforts. A major difference between the current measurement-based reform efforts and those of previous decades is that major changes in measurement formats are being introduced. Complications surface when construct validity becomes the central organizing concept. There will be questions about its limits, particularly as to the consequences of uses and interpretations of assessment results (Linn, 1994).

Student performance and achievement are related to achieving benchmarks which reflect the fluid nature of learning and evaluating, according to Larter and Sullivan (1993). Their report identifies levels of student performances for tasks which target language and mathematics. While the benchmarks are not tests, they do provide insight into the testing process.

The purpose of the "Review of Literature and Survey Results for the Utah State Core Curriculum Performance Assessment" (1992) was to describe the nature, design, and use of performance assessments across the United States. Crucial issues identified were that tasks ought to test key concepts for each subject and "require students to think critically" (p.7).

Herman (1992) discusses criteria upon which to gauge the quality of performance assessments, which include (a) consequences, (b) fairness, transfer and generalizability, (c) content quality, (d) content coverage, (e) meaningfulness, and (f) cost and efficiency. Mississippi

educators should insure that test costs correlate with meaningfulness and that accountability rests upon carefully considered criteria.

The nature of PBA tests and their design must be considered when discussing potential usefulness. Yen (1993) discusses the issue of local item dependence (the answer to one item is dependent upon the answer to another item). Yen maintains that performance assessments are more likely to produce local item dependence than regular multiple choice tests; and item dependence is one of the problems of performance assessments that still must be resolved.

Purpose of the Study

The purpose of this study was to examine the attitudes of students, teachers, and administrators following the performance based assessment initiated in Mississippi during the fall of 1994. Participants were randomly selected to respond to the investigators' survey instrument which was distributed to fourteen south Mississippi schools in the interval between the statewide pilot assessment and the schools' receipt of results. The intervening period likely represented the "best read" of participant attitudes, uncontaminated by test performance data. Research questions guiding the study were as follows:

1) Were initial responses to Performance Based Assessment across administrators, teachers, and students positive?

2) Was grade level response across teachers and students correlated?

3) Were teachers' expectations of changing teaching methods and administrator expectation for altering the curriculum in response to test data correlated?

The investigators hypothesized that Mississippi teachers and students at elementary, junior high, or senior high sites would report similar attitudes toward the new assessment procedure and instrument. Responses were expected to differ among informants at the various grade levels.

Additionally, it was hypothesized that administrators and teachers would support formative uses of the test data. With evidence of their combined interest in

formative uses, the investigators might hope that test data might actually be put to use in redesigning instructional delivery in Mississippi classrooms.

A survey instrument devised by investigators was used to compile information about attitudes and perceptions of participants in the fall testing. Three forms of the instrument addressed related questions for participants--administrators, teachers, and students. The surveys identified attitudes, perceptions, and expectations surrounding participation in an alternative assessment experience.

The student survey was designed to accommodate children in grades one through ten. Students responded to ten questions, including: testing period tolerance, difficulty of test, understanding of test directions, understanding of test items, persistence of test taker, provision for bathroom breaks, interaction with teachers, use of test aids, and format preference for PBA over achievement (scantron) testing, and perception of the performance based test.

Five of the student survey questions paralleled teacher survey questions. In addition to the paralleled items, teachers were asked about the use of the performance assessment results as a diagnostic tool, expectations for classroom instruction changes, and if knowledge required for this assessment appeared appropriate for grade level taught, including special service students. Five of the teacher survey questions correlated with questions asked administrators. Additional questions posed to administrators pertained to preparation of teachers administering the test, expected effect of test results on goals and objectives of school, and overall impression of the newly used assessment program.

Prior to the administration of the survey, contact persons agreed to seek responses from 10% of the accessible population. The accessible population consisted of personnel and students participating in the Fall 1994 Mississippi performance-based assessment at fourteen school sites. All surveys were collected and encoded prior to the statewide release of performance based assessment scores.

The analysis of administrator, teacher, and student reaction and attitudes to the fall Mississippi performance based assessment included computation of descriptive and correlational statistics for four dimensions. The four dimensions are: 1) perception of the instrument; 2) materials and teacher support during test; 3) physical environment at test site; and 4) persistence by individuals.

Five questions relate to students' perceptions of the performance testing. Students surveyed, on the average record a moderate response to the question of test difficulty (mean=1.92). Respondents relate that they understood the test instructions (mean=2.52). Significant differences are found for grades 1 through three as compared to the remaining grades on format preference questions. The younger students obviously prefer showing their work rather than completing a multiple-choice format; however, the mean across all grades for both questions is slightly positive (2.13). Students understand the test questions, (mean=2.39) however, significant differences surface between grade levels. The eighth and ninth grades responded less positively to test questions. For older students the questions apparently align less well with their actualized curriculum.

Two questions concern the use of materials and teacher interaction during the test. A moderately high response for interaction with teacher (mean=2.48) indicates a uniform appreciation for teacher interaction. Respondents in most grades report using the test aids (mean=2.48); however, ninth graders (mean=2.04) are significantly less likely to use test aids.

Two questions query the test environment. Students clearly felt bathroom breaks are inadequate (mean=1.80). Students indicate the daily test period may be too long (mean-1.95). Students in some grades report significantly different perceptions on this question.

A tenth question concerns the persistence of students on the test. They report persisting, trying hard, when taking the test (mean=2.87). This finding is consistent across grades.

Student differences are clearly seen on the graph; differences appear especially notable in some middle grades as compared to lower elementary grades. Did some middle school students (and their teachers) perceive the PBA differently? Was the fall assessment more like *high stakes* to participants in grades targeted for 1995 state evaluations of districts? The differences are troublesome enough for concerned educators to question the impact of accountability when considering perception and test data for middle grades. Does pressure to perform inversely affect both perception and test scores? The other possibility is that students in these grades were educationally less prepared for the testing combination than students in earlier or later grades. However, the two grades with the lowest attitude measure (grades 6 and 8, see Graph) have the highest national percentile ranks in the target grades (45%), according to the Mississippi State Department of Education. Given this indicator, we as teachers think it likely that some sort of perception of being at risk influenced survey and test data for at least some of these grades.

Teachers (N = 58) see the test as having just slightly more Difficulty than their students record (teacher mean 1.79 compared to student mean 1.92). They perceive Student Understanding of Directions at the same level recorded by their students (teacher mean 2.51, compared to student mean 2.52). The teachers record slightly lower, but still effective, Interaction with Teacher during test (teacher mean 2.21 compared to student mean 2.48). Teachers' perceptions of Tolerance for Testing Period is more positive than students' perception (teacher mean 2.13 compared to student mean 1.95). On Understanding Questions, teachers indicate less student understanding than do their pupils (teacher mean 1.79 compared to student mean 2.39).

Administrators from 11 of the 14 schools returned their surveys. For the most part their responses are positive (mean 2.28). Their perceptions of performance testing and its diagnostic uses indicate some misalignment between state assessment goals and objectives and

student performance and diagnostic uses of the test results differ in that diagnostic possibilities are positively perceived (mean 2.5) and student performance expectations are negatively perceived (mean 1.27).

The outcomes of this study confirm that initial responses to a new type of assessment were favorable in the period preceding return of performance data. Younger students respond most favorable to performance based assessments, middle school and older students are less comfortable with the process (see Graph). Did comfort level affect performance? Though no clear connection is inferred, Mississippi students (grades 4-9) at the initial testing are performing in the low-average range on seven out of twelve sections in language and math; they are at a high-average on the remaining five sections. Grade four performance is in the high-average range for both language and math, but next to lowest for NPR (40%) Fourth graders (n=75) record the second highest attitudinal response (22.33; range 10-30) to performance testing of grades 4-10. During next year's testing, we need to examine the relationship between attitude and performance scores, especially for targeted grades.

As teachers and administrators, we hope that Mississippi considers the test as a formative evaluation of current practices and provides teachers and their districts with more of the excellent materials generated prior to the assessment. *Hattiesburg American* reporter Thad Slaton, February 6, questioned district board members, students, and administrators about the tests. In a representative statement, John Frisk, Hattiesburg, predicts, "We'll be looking at specifics to determine what kind of things we can do to improve." (Slaton, 1995)

Performance data was omitted in early newspaper reports (i.e. *Clarion Ledger*, February 5, 1995); later reports included criterion results along with percentages for the norm-based portion of the test package. News reports compared districts on the achievement portion of the Riverside Integrated Assessment to each other and to districts' former Stanford Achievement Test levels. Mississippi uses achievement

scores to assign accreditation levels (1-5; worst to best) and to compare district against each other. Though the fall testing was presented to participants as an exploratory testing, and the state did not assign levels immediately, these scores will reportedly figure in with the October 1995 results (Hayden, 1995). The issues of state-level accountability counter point possible formative, teacher-based uses of test data.

Education in Mississippi can benefit from other states' exploration of performance based assessment. Goals are usually not met immediately, rather require patience, moderate expectations, and ongoing evaluation. Basic goals (measuring student performance through testing batteries and instructional improvement) require years to implement. What schools do with their special education students is by no means consistent across districts; under pressure schools will likely exclude some numbers of their low achieving students (Zlatos, 1995). In the future we will need to look at the apples and oranges dilemma, districts which include and are subsequently compared to districts and states which exclude low achieving students. In *Educational Digest* (1995), the managing editor Ken Schroder reminds educators that we need to test the test and testing procedures. School, district, and state department personnel need to examine all aspects of the new testing procedures.

Action Plan

As educators we suggest the following plan of action. The following suggestions make sense in terms of performance assessment; they also represent higher forms of learning.

- ◊ Discussion with students will change from grade assignment to meeting and exceeding criteria.
- ◊ Teachers will move toward embedding skills in context, crossing discipline areas.
- ◊ Teachers will recognize the many genres in which students can perform (i.e. writing essays or making graphs).
- ◊ Teachers will expect students to justify answers.

Although the initial response to performance assessment is positive, problems exist in the transition period between purely objective testing to multiple measures, including performance evaluation. Two problems exist at the classroom level: students have not developed the ability to explain the procedures by which they derive their answers, nor can they formulate open-ended problems. Both obstacles can be addressed through metacognitive discussion among teachers and their students. In the remaining months of the 1994-1995 school year, teachers can take action and prepare their students for increased performance in the fall of 1995. Teachers, administrators, and the State Department of Education need to align with the objectives of the Iowa testing package. A meaningful relationship must be orchestrated between national curriculum expectations and state guidelines. The local school districts realize they have curricula autonomy, but at a price. The more districts diverge from the national standards, the more their students may be penalized on standardized performance based or norm-based testing. Perhaps we need to examine a 60% match between national subject objectives and state mandated curriculum structure. This would leave 40% of curriculum package available for differentiation at the local level.

References

Garten, T. R., J. A. Hudson, and H. A. Gossen. Teacher induction using a shared lesson design model. *NASSP Bulletin*, 77 (1993): 76-81.

Hayden, C. (1995). Schools try new standardized test; high, low scores remain same. The *Clarion Ledger*, Sunday, February 5, 1995.

Herman, J. L. *Accountability and Alternative Assessment: Research and Development Issues*. ERIC, 1992. ED 357 037.

Hymes, D. *Making sense of testing and assessment*. ERIC, 1993 ED 363 643.

Koretz, D. (1992). *Can portfolios assess student performance and influence instruction?* The 1991-92

Vermont Experience. SCE Technical Report 371.

Larter, S., and M. Sullivan. Toronto's Elementary Benchmark Program. ERIC, 1993. ED 363 644.

Latting, J. *Assessment in Education: A Search for Clarity in the Growing Debate*. ERIC, 1992. ED 363 360.

Linn, R. L. Performance Assessment: Policy promises and technical measurement standards. Educational Researcher, 23(1994): 4-14.

Mississippi State Department of Education. Overview of the Performance Assessments for the Iowa Tests of Basic Skills (ITBS) and the Tests of Achievement and Proficiency (TAP). Jackson, MS: 1994.

PROFILES Corporation. *Review of Literature and Survey Results for the Utah State Core Curriculum Performance Assessment*. ERIC, 1992. ED 366 649.

Rafferty, E. A. *Urban Teachers Rate Maryland's New Performance Assessments*. ERIC, 1993. ED 358 168.

Slaton, T. (1995). Students find tests tougher than normal. *Hattiesburg American*, Monday, February 6, 1995.

Taylor, R. D. *Reassessing performance based assessment*. ERIC, 1993. ED 366 647.

Yen, W. M. Scaling performance assessments: strategies for managing local item dependence. Journal of Educational Measurement, 30(1993): 187-213.

Zlaltos, B. (1995). Don't test, don't tell; Inflates test scores. Summary in Educational Digest, February, 1995. Originally appeared Nov. 1994 in the *American School Board Journal*.

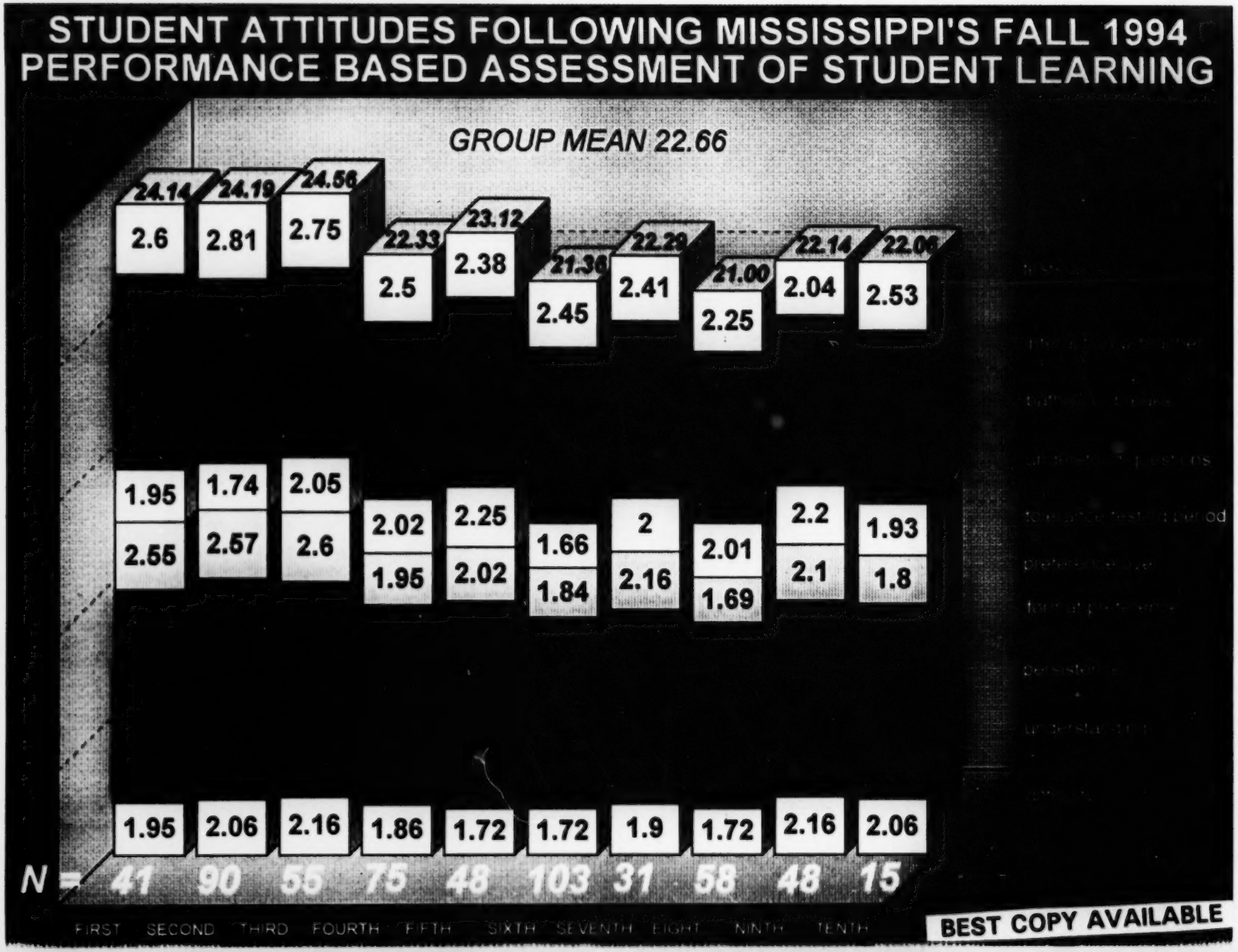
Project organizers for this study include: Susan Batson, Denise Boyd, Ann Butler, Larry Callahan, Georgia Field, Cindy Garner, Mary Alice Jackson, Marie Keyes, Christy McLemore, Glenda Nickson, Charlotte Norris, Connie

Parish, Gale Parish, Ginger Thornton, and Linda Walters.

Dr. Read M. Diket is Assistant Professor of Art in Education and Director of the Honor's Program and Creative Scholars Center at William Carey College.

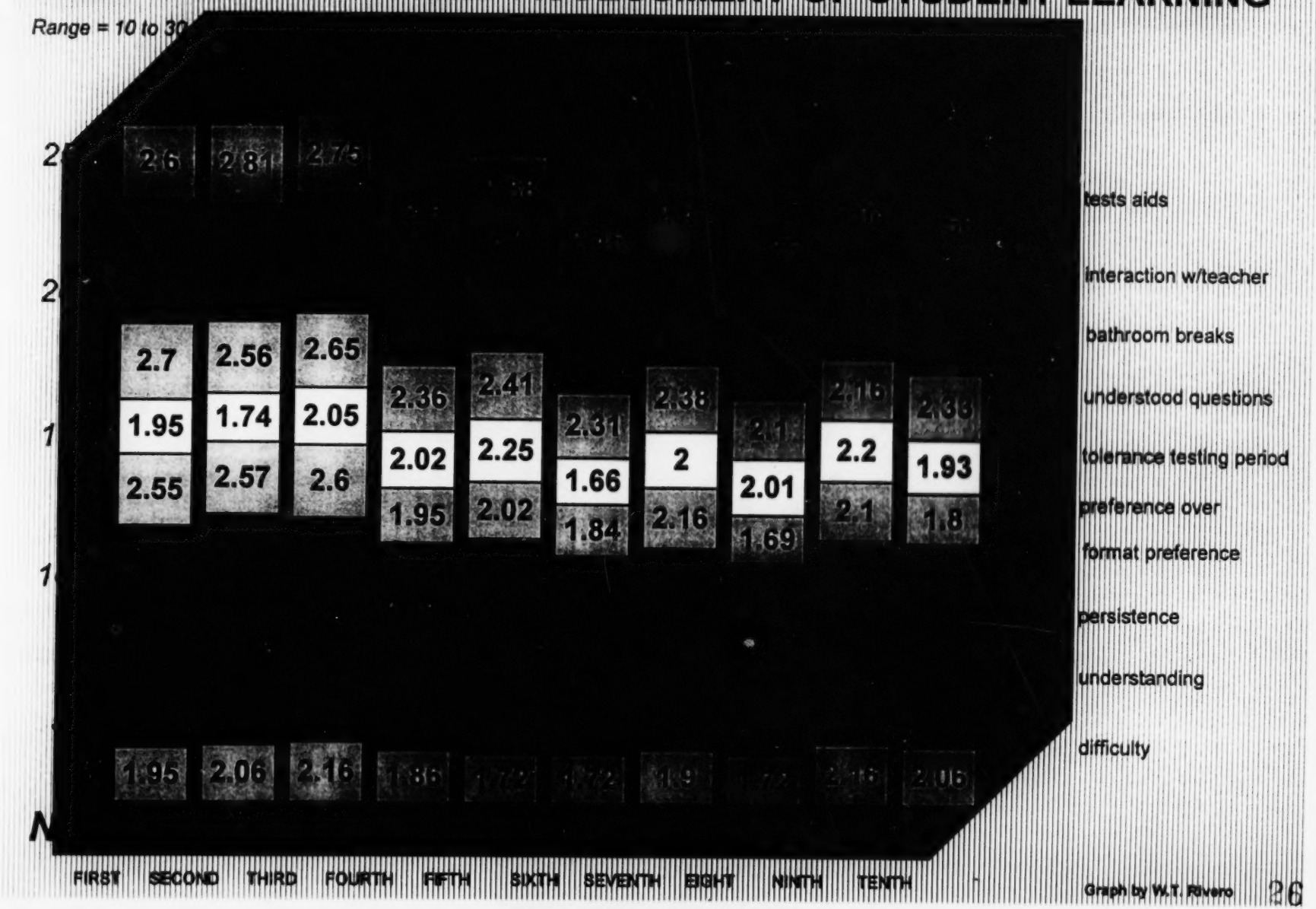
Felicia Robinson is a graduate student at William Carey College and a teacher at Mendenhall High School.

STUDENT ATTITUDES FOLLOWING MISSISSIPPI'S FALL 1994 PERFORMANCE BASED ASSESSMENT OF STUDENT LEARNING



BEST COPY AVAILABLE

STUDENT ATTITUDES FOLLOWING MISSISSIPPI'S FALL 1994 PERFORMANCE BASED ASSESSMENT OF STUDENT LEARNING



BEST COPY AVAILABLE